PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	External Validation of the 4C Mortality Score for Hospitalized COVID-19 Patients in the RECOVER network
AUTHORS	Gordon, Alexandra June; Govindarajan, Prasanthi; Bennett, Christopher L.; Matheson, Loretta; Kohn, Michael; Camargo, Carlos; Kline, Jeffrey

VERSION 1 – REVIEW

REVIEWER	Hohl, Corinne M.
REVIEW RETURNED	13-Sep-2021

GENERAL COMMENTS	I commend the authors for this study. Please consider the following: 1. Wynants et al.'s systematic review is quite dated, given the speed of publication on COVID-19. I recommend updating your search, as more high-quality scores have been published that should be discussed in this paper. 2. There is no mention of consecutive eligible enrolment of patients in this dataset. This is leaves this external validation study open to selection bias, and may have influenced who got enrolled at the site level (with sicker patients more likely to be enrolled). This probably explains the high reported mortality, which exceeds that of other registries that have enrolled consecutive eligible patients without selection bias from the early pandemic. This would have influenced the performance characteristics of the rule. 3. Palliative patients were not excluded from this study. The finding that nursing home patients commonly died is therefore not surprising. As a result, this study suffers from self-fulfilling prophecy bias, whereby the code status of the patient and treatment provided as a result of the code status likely contributed to the patient's death. These patients would have needed to have been excluded for this clinical decision rule (in both derivation and validation, and in this external validation) to be truly useful for clinical decision-making. Physicians need a prediction rule among non-palliative patients to understand in whom intubation and resuscitation will be futile. As developed and validated the rule risks predicting mortality among those who received limited care due to their palliative code status, which is a self-fulfilling prophecy. 4. The dataset used is from before September 2020 and may not reflect circulating variants or the use of modern therapies for severe COVID-19. A breakdown of patient enrolment over time would be useful to understand how many were treated with modern COVID-19 therapies to understand its applicability to current times.

	5. Was there a maximal proportion of data that could be imputed in this study? 6. The finding that nursing home residence was associated with mortality is undoubtedly true, but likely the author's estimate is confounded by not excluding palliative patients. A sensitivity analysis could be conducted to understand whether the rule's performance is still robust after excluding palliative patients, and to understand whether the effect size of nursing home residence is robust. Alternatively, all patients coming from nursing homes could be excluded in a sensitivity analysis to see if the rule's performance remains robust to explore the possibility fo self-fulfilling prophecy bias, if code status was not collected in this dataset. 7. Please expand the limitations section with the above considerations.

REVIEWER	Halasz, Geza Guglielmo da Saliceto Hospital
REVIEW RETURNED	13-Sep-2021

GENERAL COMMENTS	In the present study, Gordon and coworkers externally validated the 4c mortality score on 7961 patients hospitalized for covid-19 pneumonia belonging to RECOVER registry. In brief, the authors showed that 30-day mortality was increased with age 80+ years ,male sex and nursing home/assisted living facility residence.The 4C Score had comparable discrimination in the RECOVER dataset compared with the original 4C validation dataset. Although not really novel the study is well written and confirmed several previous findings about risk factors for COVID-19 mortality. The following suggestions/remarks are aimed at further clarification of some issues in the manuscript. Major comments • The study lacks important laboratory parameter such as D-dimer or troponin which strongly affect the patient's prognosis • Among the presented laboratory values not only CRP but also creatinine and Lymphocyte count are lacking for a half of patients • As for smoking also the nursing home/assisted living facility should be adjusted for other variables like age and comorbities Minor Comments • It would be better to further characterize cardiovascular disease, including the percentage of patients suffering from ischemic heart disease. • I think that the estimation of mortality risk in patients with COVID-19 pneumonia is of paramount importance. In this context, this new machine learning based score has demonstrated a comparable accuracy with respect to the 4c mortality score and could be mentioned in the discussion.(Halasz G, Sperti M, Villani M, et al. A Machine Learning Approach for Mortality Prediction in COVID-19 Pneumonia: Development and Evaluation of the Piacenza Score. J Med Internet Res. 2021 May 31;23(5):e29058. doi: 10.2196/29058. PMID: 33999838; PMCID: PMC8168638.)

VERSION 1 – AUTHOR RESPONSE

Reviewers' Comments to the Author: Reviewer: Corinne Hohl	Author Responses
Wynants et al.'s systematic review is quite dated, given the speed of publication on COVID-19. I recommend updating your search, as more high-quality scores have been published that should be discussed in this paper.	We recognize that the COVID-19 literature has progressed rapidly and that there have been other scoring models proposed and discussed since we initiated our study. We have added additional references (see below) to our manuscript discussion to better place our work within the context of other models. It should be noted, however, that at the time of our study during the early pandemic, the 4C Mortality Score was among the most promising. All models have potential biases, and evaluation and recalibration of existing models provides valuable information for providers. In addition, the Wynants et al. systematic review has been recently updated (https://doi.org/10.1136/bmj.n236). The authors express the same concerns and again identify the 4C model as a promising prognostic model that should be validated by other studies. *Update to living systematic review on prediction models for diagnosis and prognosis of covid-19. BMJ. 2021 Feb 3;372:n236.* *Gupta RK, Marks M, Samuels THA, Luintel A, Rampling T, Chowdhury H, et al. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study. Eur Respir J [Internet]. 2020 Dec 24 [cited 2021 Jun 7];56(6). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7518075/* *. Cho S-Y, Park S-S, Song M-K, Bae YY, Lee D-G, Kim D-W. Prognosis Score System to Predict Survival for COVID-19 Cases: a Korean Nationwide Cohort Study. J Med Internet Res. 2021 Feb 22;23(2):e26257.*

	Berenguer J, Borobia AM, Ryan P, Rodríguez-Baño J, Bellón JM, Jarrín I, et al. Development and validation of a prediction model for 30-day mortality in hospitalised patients with COVID-19: the COVID19 SEIMC score. Thorax. 2021 Feb 25; *Nicholson CJ, Wooster L, Sigurslid HH, Li RH, Jiang W, Tian W, et al. Estimating risk of mechanical ventilation and in-hospital mortality among adult COVID-19 patients admitted to Mass General Brigham: The VICE and DICE scores. EClinicalMedicine. 2021 Mar;33:100765.* *Magro B, Zuccaro V, Novelli L, Zileri L, Celsa C, Raimondi F, et al. Predicting in-hospital mortality from Coronavirus Disease 2019: A simple validated app for clinical use. PLoS One. 2021;16(1):e0245281.* *King Jr JT, Yoon JS, Rentsch CT, Tate JP, Park LS, Kidwai-Khan F, et al. Development and validation of a 30-day mortality index based on pre-existing medical administrative data from 13,323 COVID-19 patients: The Veterans Health Administration COVID-19 (VACO) Index. PLoS One. 2020 Nov 11;15(11):e0241825.* *Goodacre S, Thomas B, Sutton L, Burnsall M, Lee E, Bradburn M, et al. Derivation and validation of a clinical severity score for acutely ill adults with suspected COVID* *-19: The PRIEST observational cohort study. PLoS One. 2021;16(1):e0245840.* *Gupta RK, Harrison EM, Ho A, Docherty AB, Knight SR, Smeden M van, et al. Development and validation of the ISARIC 4C Deterioration model for adults hospitalised with COVID-19: a prospective cohort study. Lancet Respir Med. 2021 Apr 1;9(4):349–59.*

	https://www.medrxiv.org/content/10.1101/2021.07.28.21261283v1
There is no mention of consecutive eligible enrolment of patients in this dataset. This leaves this external validation study open to selection bias and may have influenced who got enrolled at the site level (with sicker patients more likely to be enrolled). This probably explains the high reported mortality, which exceeds that of other registries that have enrolled consecutive eligible patients without selection bias from the early pandemic. This would have influenced the performance characteristics of the rule.	Sites reviewed all subjects who met the criteria for inclusion and subsequently enrolled patients until they reached their quota. Some sites did enroll more than others, as different hospitals had different quotas based on size and patient load. This nonprobability sampling technique is consistent with consecutive enrollment.
Palliative patients were not excluded from this study. The finding that nursing home patients commonly died is therefore not surprising. As a result, this study suffers from self-fulfilling prophecy bias, whereby the code status of the patient and treatment provided as a result of the code status likely contributed to the patient's death. These patients would have needed to have been excluded for this clinical decision rule (in both derivation and validation, and in this external validation) to be truly useful for clinical decision-making. Physicians need a prediction rule among non-palliative patients to understand in whom intubation and resuscitation will be futile. As developed and validated the rule risks predicting mortality among those who received limited care due to their palliative code status, which is a self-fulfilling prophecy.	There were only 133 palliative care patients in the original dataset of 7961, so 1.67%. However, we understand that the reviewer means for us to exclude palliative care, do not intubate, and do not resuscitate. In other words, include only "full code" patients. In that case, our dataset decreases by 1159 (14.6%) to 6802. We have re-run the analysis and re-created the tables using the dataset of 6802 full-code patients.
The dataset used is from before September 2020 and may not reflect circulating variants or the use of modern therapies for severe COVID-19. A breakdown of patient enrolment over time would be useful to understand how many were treated with	In this study, most of the patients were enrolled in March and April of 2020. The major practice change that affected mortality was the RECOVERY trial, which was published in February 2021 and resulted in the widespread use of dexamethasone in patients with COVID-19. The enrollment dates for that trial were similar to our study. However, we agree that

modern COVID-19 therapies to understand its applicability to current times.	practice changes, use of other COVID-19 therapies, and other epidemiological changes could affect the applicability of the model. We have added this to the manuscript as a limitation.
Was there a maximal proportion of data that could be imputed in this study?	In the revised manuscript, for bronchiectasis and pulmonary fibrosis, we assumed blank meant not present We did not set a maximal proportion of data to impute, but the proportions missing were < 1.75% except for smoking (12.7%), total bilirubin (10.0%), BMI (10.9%), and CRP (38.8%). Missing CRP is an acknowledged limitation and we evaluated the 4C model both with and without CRP.
The finding that nursing home residence was associated with mortality is undoubtedly true, but likely the author's estimate is confounded by not excluding palliative patients. A sensitivity analysis could be conducted to understand whether the rule's performance is still robust after excluding palliative patients, and to understand whether the effect size of nursing home residence is robust. Alternatively, all patients coming from nursing homes could be excluded in a sensitivity analysis to see if the rule's performance remains robust to explore the possibility for self-fulfilling prophecy bias, if code status was not collected in this dataset.	As above, we re-did the analysis including only full-code patients.
Please expand the limitations section with the above considerations.	We have revised the limitations section to address additional concerns.

Reviewer: 2	Author Responses

Dr. Geza Halasz, Guglielmo da Saliceto Hospital	
Comments to the Author: In the present study, Gordon and coworkers externally validated the 4c mortality score on 7961 patients hospitalized for covid-19 pneumonia belonging to RECOVER registry. In brief, the authors showed that 30-day mortality was increased with age 80+ years, male sex and nursing home/assisted living facility residence. The 4C Score had comparable discrimination in the RECOVER dataset compared with the original 4C validation dataset. Although not novel the study is well written and confirmed several previous findings about risk factors for COVID-19 mortality. The following suggestions/remarks are aimed at further clarification of some issues in the manuscript.	
Major comments • The study lacks important laboratory parameter such as D-dimer or troponin which strongly affect the patient's prognosis	It is not standard practice to get D-Dimer and Troponin in the Emergency Department for COVID positive patients in the US. These laboratory parameters were not included in the RECOVER Network's dataset. In addition, the 4C score does not include/consider D-dimer and/troponin.
• Among the presented laboratory values not only CRP but also creatinine and Lymphocyte count are lacking for a half of patients	We have removed Lymphocyte from the table since it is not used in either the 4C score or the multivariable model. Creatinine was missing for a large number of patients, but almost all of these patients had BUN values. Kidney disease was defined as Cr >= 1.2 or BUN >= 40, and the 4C Score does not use Cr.
• As for smoking also the nursing home/assisted living facility should be adjusted for other variables like age and comorbities	Smoking and assisted living are not included in the 4C score, they are only used here in the multivariable model, which does account for other variables like age and comorbidities.
Minor Comments • It would be better to further characterize cardiovascular disease, including the	We agree that cardiovascular disease is a contributing factor to COVID-19 mortality and that there may be variation based on the type of heart disease. However, based on the limitations of a

percentage of patients suffering from ischemic heart disease.	retrospective registry, we are only able to categorize as a binary variable of yes/no to “heart disease” and are not able to obtain additional subclassification data to clarify the underlying type of heart disease. Thus, we do not have this level of detail in our dataset for this analysis.
• I think that the estimation of mortality risk in patients with COVID-19 pneumonia is of paramount importance. In this context, this new machine learning based score has demonstrated a comparable accuracy with respect to the 4c mortality score and could be mentioned in the discussion. (Halasz G, Sperti M, Villani M, et al. A Machine Learning Approach for Mortality Prediction in COVID-19 Pneumonia: Development and Evaluation of the Piacenza Score. J Med Internet Res. 2021 May 31;23(5):e29058. doi: 10.2196/29058. PMID: 33999838; PMCID: PMC8168638.)	We acknowledge that the state of COVID-19 research has progressed rapidly and that many other scoring systems have been proposed. We have added several of these to the discussion.

VERSION 2 – REVIEW

REVIEWER	Hohl, Corinne M.
REVIEW RETURNED	14-Dec-2021

GENERAL COMMENTS	The authors have addressed the concerns initially raised during peer-review. The authors validated a modified 4C score, which has inferior performance characteristics than the CCMS (4C AUROC 0.776, 95% CI 0.76-0.79 versus CCMS AUROC 092, 95% CI 0·89–0·93) in validation. The authors may wish to qualify their conclusions as this difference in performance is rather substantial. As a practicing clinician I would adopt the score with higher perfo Thus, I recommend rephrasing statements like "The Coronavirus Clinical Characterisation Consortium Mortality Score (4C Score) is the most promising COVID-19 mortality risk model" in the abstract introduction. Other than the above, I have no concerns and congratulate the authors and thank the for the additional work they have done.

REVIEWER	Halasz, Geza Guglielmo da Saliceto Hospital
REVIEW RETURNED	30-Nov-2021

GENERAL COMMENTS	The authors have addressed most of the reviewer's comments and clarified methodological limitations.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1 (Dr. Corinne M. Hohl) recommended rephrasing statements like "The Coronavirus Clinical Characterisation Consortium Mortality Score (4C Score) is the most promising COVID-19 mortality risk model". We identified it as "a promising COVID-19 mortality risk model."

We made several minor typographical changes and an address change for one of the authors. All changes are in the marked and clean copies.